基于改进最小分类误差准则算法的深度学习研究——以台风卫星云图为例 *

郑宗生，侯　倩，邹国良†，卢　奇

(上海海洋大学 信息学院，上海 201306)

摘　要：针对传统基于最小分类误差准则(MCE)建立的目标函数存在样本错分类时网络出现的梯度反向问题，引入最小分类误差准则，定义带修正项的 FMCE 目标函数。以较高精度的交叉熵作为基函数，将 FMCE 作为修正函数，提出改进交叉熵目标函数 CE-FMCE，使得网络在反向传播过程中提升标签类输出的概率。CE-FMCE 不仅克服了传统 MCE 目标函数的梯度反向问题，还弥补了交叉熵函数对非标签集梯度不作区分处理的不足。分别在自建台风云图数据集和通用数据集 MNIST 上对 CE-FMCE 和 MSE、交叉熵、MCE、M3CE 进行对比实验，实验结果表明 CE-FMCE 优于其他目标函数。

关键词：深度学习；卷积神经网络；交叉熵；最小分类误差准则；台风等级

中图分类号：TP3　　doi: 10.3969/j.issn.1001-3695.2018.04.0335

Research on deep learning based on improved minimal classification error criterion algorithm: take typhoon satellite image as example

Zheng Zongsheng, Hou Qian, Zou Guoliang†, Lu Qi

(*College of Information*, *Shanghai Ocean University*, *Shanghai* 201306, *China*)

Abstract: For the traditional objective function based on the minimum classification error criterion (MCE) , there existed the problem of gradient inversion in the network when the sample was misclassified. This paper introduces the minimum classification error criterion and defines the FMCE objective function with correction term. This paper used the higher precision cross entropy as a basis function and FMCE as a correction function, and proposes an improved cross-entropy objective function CE-FMCE, which makes the network increase the probability of label class output in the reverse propagation process. CE-FMCE not only overcomes the gradient inversion problem of the traditional MCE objective function, but also compensates for the lack of differentiation of the cross-entropy function for non-label set gradients. This paper compared CE-FMCE and MSE, cross-entropy, MCE, and M3CE on self-built typhoon image dataset and general dataset MNIST. The experimental results showed that CE-FMCE was superior to other objective functions.

Key words: deep learning; convolutional neural network; cross-entropy; minimum classification error criterion; typhoon rating

0　引言

深度学习是当前流行的机器学习算法[1]，针对浅层神经网络对复杂分类问题泛化能力弱的缺点，其通过模拟人脑进行逐层学习，提取自然信息的深层抽象特征来提高泛化能力[2]。目前，深度学习在模式识别、智能视频、机器视觉等研究领域取得了一系列成果[3~6]。卷积神经网络（convolutional neural networks, CNN）作为深度学习算法成功应用的模型之一，由 LeCun[7]在 1989 年提出。直到 2012 年在 ImageNet 评测问题中将错误率降低了 9%，CNN 才在图像识别应用方面获得重大进展[8]。目前国内外研究学者对 CNN 的权值优化方法主要集中在选择合适的网络参数（如变尺寸卷积核[9]、参数池化[10]、Dropout 置零率[11]等）；使用具有更好稀疏特性的激活函数(如 Relu、Leaky Relu、PRelu 等)。其中权值优化方法中，目标函数至关重要，目标函数作为 CNN 模型中重要的组成部分，其值越小表示模型的鲁棒性越好[12]。在网络的训练过程中它不仅能表示当前网络的状态，还在反向传播的梯度下降算法中提供参数的梯度。如果输出层的梯度太小，梯度经过深层网络衰减后，低

收稿日期：2018-04-18；**修回日期：**2018-06-14　**基金项目：**国家自然科学基金资助项目（41671431）；上海市科委地方高校能力建设项目（15590501900）

作者简介：郑宗生（1979-），男，副教授，博士，主要研究方向为海洋信息化、深度学习；侯倩（1992-），女，硕士研究生，主要研究方向为图像处理、深度学习；邹国良（1961-），男（通信作者），教授，博士，主要研究方向为海洋信息处理及应用（glzou@shou.edu.cn）；卢奇（1990-），女，硕士研究生，主要研究方向为图像处理.

层基本接受不到有效的训练信号。如何使网络权值达到最优从而提高泛化能力，其中构建目标函数成为研究的热点。

在卷积神经网络算法中，通常使用的目标函数有最小均方误差(mean square error, MSE)、交叉熵(cross entropy)等，其中 MSE 更适用于解决回归问题。Papoulis 等人[13]认为在分类问题中 MSE 估计的是后验概率，当网络的输出函数为 sigmoid 时网络存在梯度消失的问题。随着 CNN 的发展，研究证明交叉熵损失函数较 MSE 平坦区域更少，使得网络更容易跳出局部最佳值[14, 15]，所以对于多分类问题使用交叉熵作为目标函数能够取得更好的效果。因此，目前对于目标函数的研究，大多是针对特定的问题，在交叉熵函数的基础上引入与之相关的参数项。例如归喆针对人脸识别问题，增加个体内差异的损失函数作为正则项，使得网络学习到属于同一个人的特征向量在空间内尽可能相似[12]。但是交叉熵函数在梯度下降算法中并不对非标签维的梯度作区分处理，而是对它们作相同的训练。这就导致分类器不能很好的将标签类和最容易混淆的类区分，从而降低模型的正确率。

如果一个分类器在训练的过程中，能将标签类与最容易混淆的类别区分，那势必会降低误差率。Juang 等人[16]最先将最小分类误差方法(minimum classification error,MCE)引入到浅层神经网络的训练过程中，他们提出的目标函数考虑非标签集中混淆程度最大的一个类别来降低分类的错误率但是基于最小分类误差构造的 logistic 目标函数存在梯度消失的问题。当深层卷积神经网络通过反向传播算法调整网络权值时，处于最顶层的目标函数一旦出现梯度饱和，将会影响网络的训练，因此传统的 logistic 函数并不适合直接应用到深度卷积神经网络中。为了克服梯度饱和的问题，Feng 等人[17]提出基于最大间隔最小分类误差（M^{3}CE）建立目标函数，但当样本被错分类时，M^{3}CE 出现部分非标签维所对应的梯度方向与交叉熵梯度方向相反，这将导致网络在反向传播的过程中信息不能充分训练且收敛速度降低，因此限制了其推广使用。

为了克服传统的 MCE 目标函数存在的梯度方向问题，本文引入最小分类误差准则，定义带修正项的目标函数 FMCE，提出基于最小分类误差准则的修正交叉熵目标函数（CE-FMCE）。CE-FMCE 不仅克服了传统 MCE 目标函数的梯度方向问题，而且弥补了交叉熵函数对非标签维的梯度不做区分处理的不足。并且将此目标函数应用于台风卫星云图自建数据集和手写字体库 MNIST 数据集，通过在卷积神经网络框架下进行对比实验，证明本文提出的目标函数 CE-FMCE 的有效性。

1　传统卷积神经网络

卷积神经网络是由卷积层和采样层交替组成的网络拓扑结构，在网络训练过程中包括前向传播和反向传播两个过程。对于前向传播，卷积层利用本层的卷积核与输入图像进行卷积运算，实现对输入图像隐式的特征提取；采样层则对当前输入的特征图做下采样降低图像的分辨率，减少运算量从而提高网络的收敛速度。网络在反向传播中运用残差 δ 反向传播规则[18]由输出层向输入层逐层传递，使得模型目标函数不断降低，对每个神经元的权值进行更新，从而得到网络的最佳权值。两个过程循环进行，直到目标函数达到规定阈值或达到最大迭代次数为止。残差表征网络实际输出 y_{true} 与期望输出 y_{pred} 之间的误差信号，深度为 H 层的网络的输出层残差可表示为

$$\delta = \frac{\partial L(w,b)}{\partial z_H} = \nabla_{y_{true}} L \bullet \sigma'(z) \tag{1}$$

其中：$\sigma'(z)$ 表示输出层函数的导数；z_H 表示第 H 层的输入。

卷积神经网络通常采用 softmax 层作为网络的输出层函数。设 z_j 为 softmax 函数的输入，那么网络的输出层中第 j 个神经元的值表示为

$$p_j = \frac{\exp z_j}{\sum_{i=1}^{m} \exp z_i} \tag{2}$$

其中：m 表示样本类别数。

对于一个含有 N 个样本的训练集，其交叉熵函数表示为

$$L_c = -\frac{1}{N}\sum_{n=1}^{N} y_n^T \log p_n \tag{3}$$

由于网络在训练过程中的误差是所有样本误差的总和，所以为了方便分析，这里只考虑单个样本。那么输出层残差 δ_c 表示为

$$\delta_c = \frac{\partial L}{\partial z_j} = \begin{cases} p_j - 1, & j = k \\ p_j, & j \neq k \end{cases} \tag{4}$$

其中：j 表示 p 中的任意一维(即输出层中第 j 个神经元的值)；k 表示样本标签维。根据式（4）可以看出，当所求的维度是样本标签维时，即 $j = k$，此时 $\delta_c < 0$，其余维度 $\delta_c > 0$。通过分析得知，当 $j \neq k$ 时 $\delta_c = p_j$，那么交叉熵函数在反向传播过程中对所有非标签维梯度不作区分处理，而是采用相同的方式对其进行训练。这将降低网络的收敛速度且训练精度不高。

2　改进 MCE 目标函数算法及证明

2.1　改进 MCE 目标函数算法

针对传统卷积神经网络中交叉熵函数存在的问题，可以在交叉熵函数的基础上利用最小误差准则建立目标函数来优化 CNN。通过定义错分类度量，将不同的非标签维采用不同的方式进行训练，从而弥补交叉熵的不足。

基于最小分类误差定义的目标函数一般为以下三个步骤：

a)对某个类别 k，定义一个判别函数 $g_k(z)$。

b)对类别 k 中每个样本特征给定一个错分类度量 $d_k(z)$。

c)为 $d_k(z)$ 构建一个目标函数 $l_k(d_k(z))$。最终的目标函数定义如下：

$$L_{MCE}=\frac{1}{N}\sum_{n=1}^{N}l_n(d_n(z_n)) \tag{5}$$

传统方法在建立 l_k 时仅考虑 d_k 的影响，例如采用 logistic 目标函数，随着网络深度增加会出现梯度消失问题；M³CE 定义 $l_k=(1+d_k)^3$，当样本被错分类时只有错分类维梯度为正数，对比(4)式发现，这与交叉熵方向相反，这样不仅会造成收敛速度下降，而且使得信息不能充分训练。为了防止错误的信号传播到网络底层引起不可预测的错误，因此本文对 MCE 进行改进，提出目标函数 FMCE。其定义过程如下：

a)对于多分类问题网络采用 softmax 作为输出层，所以本文采用 softmax 函数作为判别函数 $g_k(z)$。

b)如果一个分类器能将正确的类别与最容易错分的类别分开，那么对网络的识别率必定有所提高。所以对错分类度量 $d_k(z)$ 的定义如式(6)所示。

$$d_k(z)=-p_k+p_r=-\frac{z_k}{\sum_{i=1}^{m}\exp z_i}+\frac{z_r}{\sum_{i=1}^{m}\exp z_i} \tag{6}$$

其中：k 是样本标签类；r 表示 softmax 函数输出中针对样本标签类最容易错判的一类。

c)当 $0<d_k<1$ 时即模型出现错误分类，传统的 MCE 目标函数会增加错误信息对网络的影响，从而降低模型鲁棒性。因此本文在 $0<d_k<1$ 时增加修正项 $\ln\sum_{i=1}^{m}\exp z_i$。所以对于单一样本 l_n 的表达式为

$$l_n=\begin{cases}d_k+\ln\sum_{i=1}^{m}\exp z_i & 0<d_k<1\\ d_k & -1<d_k\le 0\end{cases} \tag{7}$$

其中：m 表示样本类别数。

那么改进后的目标函数 FMCE 可以表示为

$$L_{FM}=\begin{cases}\frac{1}{N}\sum_{n=1}^{N}(d_k+\ln\sum_{i=1}^{m}\exp z_i) & 0<d_k<1\\ \frac{1}{N}\sum_{n=1}^{N}d_k & -1<d_k\le 0\end{cases} \tag{8}$$

2.2 改进 MCE 目标函数证明

对目标函数进行输出层残差的理论推导，证明 FMCE 的可行性。为了便于分析只考虑单一样本，即式(8)被简化为

$$L_{FM}=\begin{cases}d_k+\ln\sum_{i=1}^{m}\exp z_i & 0<d_k<1\\ d_k & -1<d_k\le 0\end{cases} \tag{9}$$

在梯度下降法中输出层残差 δ_{FM} 为

$$\delta_{FM}=\frac{\partial L_M(w,b)}{\partial z_j}=\begin{cases}\delta_2+\delta_3 & 0<d_k<1\\ \delta_2 & -1<d_k\le 0\end{cases} \tag{10}$$

其中：

$$\delta_2=\frac{\partial d_k(z)}{\partial z_j}=\begin{cases}p_j(-d_k-1), & j=k\\ -p_jd_k, & j\ne k 且 j\ne r\\ p_j(-d_k+1), & j=r\end{cases}$$

$$\delta_3=\frac{\partial(\ln\sum_{i=1}^{m}\exp z_i)}{\partial z_j}=\frac{\exp z_j}{\sum_{i=1}^{m}\exp z_j}=p_j$$

因为残差 δ_{FM} 与 d_k 的大小有关，所以进行分开讨论。

当 $0<d_k<1$ 时

$$\delta_{FM}=\delta_2+\delta_3=\frac{\partial d_k(z)}{\partial z_j}+\frac{\partial(\ln\sum_{i=1}^{m}\exp z_i)}{\partial z_j}=\begin{cases}-p_jd_k, & j=k\\ p_j(-d_k+1), & j\ne k 且 j\ne r\\ p_j(-d_k+2), & j=r\end{cases} \tag{11}$$

当 $-1<d_k\le 0$ 时

$$\delta_{FM}=\delta_2=\frac{\partial d_k(z)}{\partial z_j}=\begin{cases}p_j(-d_k-1), & j=k\\ -p_jd_k, & j\ne k 且 j\ne r\\ p_j(-d_k+1), & j=r\end{cases} \tag{12}$$

式(11)~(12) δ_{FM} 的方向与 d_k 的关系可以归纳总结为表 1 所示。

表 1　目标函数梯度分析

δ_{FM}	$0<d_k<1$	$-1<d_k\le 0$
$j=k$	-	-
$j\ne k$ 且 $j\ne r$	+	+
$j=r$	+	+

因为 softmax 的输出可以表示后验概率，所以 $p_k\in[0,1], p_r\in[0,1]$，那么 $d_k\in[-1,1]$，即当一个样本从错分类变为正确分类时，d_k 从 1 变为-1。分析表 1、式（11）（12）和(4)，发现当 $0<d_k<1$ 即样本被错分类时，本文实现了对非标签维梯度作区分处理，并且保证 δ_{FM} 梯度方向与 δ_C 一致；当 $-1<d_k\le 0$ 即样本被正确分类时，δ_{FM} 梯度的方向也与交叉熵一致。因此本文方法保证提升正确类输出的同时对非标签类作不同的训练，从而证明了本文提出的方法在理论层面的可行性。综上所述，δ_{FM} 可以作为 δ_C 的补充项，即 FMCE 可以作为交叉熵的补充来优化卷积神经网络。

因此在交叉熵的基础上引入 FMCE 构成目标函数 CE-FMCE，最终表达式为

$$L = L_C + \alpha L_{FM} \tag{13}$$

其中：$\alpha = 1/T$，T 表示一个批次中样本的数量，根据实验过程中样本的大小进行取值；$L_C = -\frac{1}{N}\sum_{n=1}^{N}\log p_{yn}$。因此 CE-FMCE 可以表示为

$$L = -\frac{1}{N}\sum_{n=1}^{N} y_n^T \log p_n + \alpha \begin{cases} \frac{1}{N}\sum_{n=1}^{N}(d_k + \ln\sum_{i=1}^{m}\exp z_i) & 0 < d_k < 1 \\ \frac{1}{N}\sum_{n=1}^{N} d_k & -1 < d_k \le 0 \end{cases} \tag{14}$$

2.3　网络结构

本文构建的卷积神经网络如图 1 所示。采用多层卷积层代替传统的卷积层，增强模型特征提取能力的同时降低卷积过程的计算量。通过权值共享和网络结构重组对 6 层神经网络进行特征学习，同时使得整个网络可以通过反向传播算法得以很好地训练并用于分类。构建模型时采用三个卷积层(C1-C3)、一个采样层(S1)和两个全连接层（F4-F5），其中 C1 和 C2 采用 32 个 5*5 大小的卷积滤波器，步长为 1；C3 采用 32 个 2*2 大小的卷积核，步长同样设为 1；采样层则采用大小为 2*2 的核对卷积后的特征图进行最大下采样，降低图像的分辨率从而提高网络的收敛速度。优化模型时选择使用 Relu 函数作为激活函数；在卷积层加入批归一化层来加速网络的训练，此外在全连接层使用 Dropout 方法来抑制模型过拟合；采用本文提出的 CE-FMCE 目标函数，根据样本数据集大小采用合适的参数 α，在反向传播过程中使目标函数最小化从而优化网络参数。

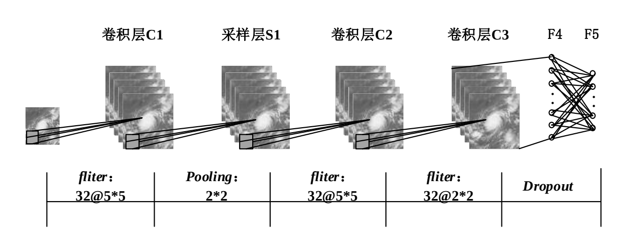


图 1　模型结构

3　实验结果及分析

本文所有对比实验均是基于 Windows 10 CPU Intel Core i5-6500M 3.2 GHz 内存为 4 GB，使用基于 Tensorflow 的 keras 深度学习框架。实验部分分为两组：为了证明提出的目标函数 CE-FMCE 对台风等级分类的可行性，第一组实验使用自建的台风卫星云图数据集，其中批尺寸为 40,则 α=0.025；为了验证 CE-FMCE 的普适性，第二组实验数据集采用通用数据集手写字体库 MNIST，此时 α=0.01，框架则与第一组实验相同都采用 2.3 节中的网络模型。此外，本文的两组对比实验中都采用 CE-FMCE 和目前被广泛使用的目标函数作对比，实验结果取得了预期的效果。

模型的精确度通过正确率检验，假设在 N 个样本中第 n 个样本的真实类标签为 y_{true}，预测类标签为 y_{pred}，则模型分类的正确率为

$$ACC(y_{true}, y_{pred}) = \frac{1}{N}\sum_{n=1}^{N} 1(y_{true} = y_{pred}) \tag{15}$$

预测得到的类标签与真正的标签集越吻合，模型分类的正确率越大。

3.1　基于 CE-FMCE 的台风等级分类

3.1.1 数据集构建

实验数据采用日本国立情报学研究所（National Institute of Informatics，NII）提供的西北太平洋 1978－2016 年“Himawari1-8”卫星观测到的高时间分辨率卫星云图资料及对应的台风强度信息。数据集包含近 1 000 多个台风过程。为了获取全天候的数据，实验采用了红外云图作为数据样本，依据日本气象台提供的台风等级标准制定 4 类台风等级标签，如表 2 所示。

表 2　台风等级标准

台风等级	最大风速/kt	最大风速/m/s
热带低气压	<34	<17
台风	≥ 34～<64	≥ 17～<33
强台风	≥ 64～<85	≥ 33～<44
超强台风	≥ 85<105	≥ 44～<54

本文将采集到的红外云图首先采用中值滤波器去除云图斑块中的噪声，有效保留图像中的边缘信息；其次采用插值缩放将云图分辨率转为 24*24；最后根据表 2 制定数据集标签，并且将台风云图转换为 24*24*1 的格式作为模型的输入。最终构建了 4 000 训练样本、800 测试样本的数据集，热带低气压、台风、强台风、超强台风训练集各 1 000 个、测试集各 200 个，部分卫星云图样本如图 2 所示。

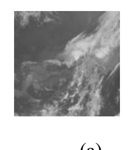
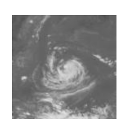
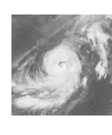
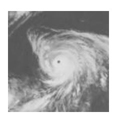

(a)

(b)

(c)

(d)

图 2　部分气象卫星云图样本（a、 b 、c、d 分别代表热带低气压、台风、强台风、超强台风）

3.1.2 结果对比

第一组实验中，将本文提出的目标函数与均方差、交叉熵函数进行对比，对模型迭代 200 次网络的正确率如图 3 所示。

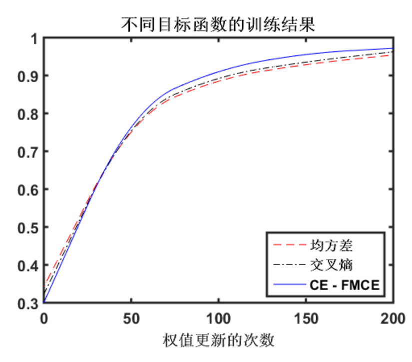


图 3　不同目标函数训练结果

对比图 3 可以发现，在现有样本的基础上，交叉熵目标函

数对网络的优化效果较均方差有一定的提高，在迭代 200 次网络的正确率为 96.25%。CE-FMCE 较其他的目标函数能取得更好的效果，网络正确率达到 98.0%，并且 CE-FMCE 收敛速度最快，均方差和交叉熵的收敛效果表现都不理想。

将本文提出的目标函数 CE-FMCE 与目前被广泛使用的多分类目标函数在训练集和测试集做对比实验，结果如表 3 所示。

表 3　不同目标函数的模型正确率

目标函数	训练集/%	测试集/%
MCE	92	78.12
均方差	95.4	82.0
交叉熵	96.25	83.38
M^3CE[17]	97.0	84.59
CE-FMCE	98.0	86.78

对比表 3 发现均方差目标函数在训练集和测试集的正确率分别为 95.4%和 82.0%；交叉熵目标函数在测试集的准确达到 83.38%；文献[17]设计的 M^3CE 目标函数在训练集和测试集的正确率都有一定的提升，训练集正确率为 97.0%，测试集正确率达到 84.59%，比交叉熵目标函数提升 1.21%。本文设计的目标函数 CE-FMCE，在模型迭代 200 次后训练集正确率为 98.0%，测试集正确率达到 86.78%，比交叉熵函数在训练集高出 1.75%，测试集高出 3.4%。CE-FMCE 在理论上修正了文献[17]提出的 M^3CE 在错分类时出现梯度反向的问题，表 3 也很好地验证了本文的改进，相比于 M^3CE 测试集的正确率提高了 2.19%，这也充分证明了 CE-FMCE 的可行性。

3.2　基于 CE-FMCE 的 MNIST 数据集对比实验

为了验证本文提出的目标函数 CE-FMCE 的普适性，第二组实验数据采用通用数据集手写字体库 MNIST。MNIST 数据集包含 0~9 共 10 类手写数字，如图 4 所示。其中包含 42 000 个训练样本，10 000 个测试样本，图像分辨率为 28*28，部分样本数据如图 4 所示。

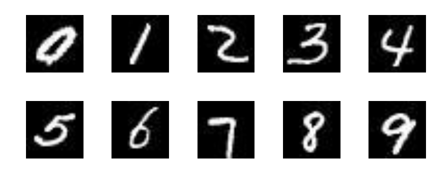


图 4　MNIST 数据集部分样本

第二组实验将 CE-FMCE 与其他的目标函数在 MNIST 数据集进行验证。网络采用与第一组实验相同的模型，将 MNIST 数据集中的样本转换为 28*28*1 的格式作为网络的输入，对模型设置迭代次数 100 次，实验结果如表 4 所示。

表 4　不同目标函数的模型正确率

目标函数	训练集/%	测试集/%
MCE	99.69	98.90
均方差	99.25	98.35
交叉熵	99.79	98.99
M^3CE[17]	99.86	99.06
CE-FMCE	99.92	99.11

从表 4 可以看出，CE-FMCE 的训练集正确率为 99.92%，测试集正确率达到 99.11%，较 M^3CE 测试集正确率提升了 0.05%，且表现能力优于其他目标函数，但是相对于其他的多分类目标函数，模型的正确率的提升并不大。分析原因在于 MNIST 数据集的数字噪声干扰少、识别难度低，所以基准正确率比较高，提升不太明显。

对比表 3 和 4 发现，CE-FMCE 对特征复杂的图像样本，模型正确率提升效果明显。

4　结束语

深度学习通过构建深层非线性网络实现复杂函数的逼近来表达自然图像的深层抽象特征，克服传统浅层神经网络对复杂分类问题泛化能力弱的问题。目前针对深度学习网络的研究大多集中在优化网络参数，目标函数作为深度学习算法的重要组成部分不仅表征当前网络的状态，还在网络的梯度下降法中提供参数梯度。因此，本文对目标函数展开研究，针对传统基于 MCE 建立的目标函数在网络中存在的梯度问题，提出基于 MCE 的修正交叉熵目标函数 CE-FMCE；并且在自建台风云图数据集和通用数据集手写字体库 MNIST，将 CE-FMCE 与目前被广泛应用于多分类问题的目标函数 MSE、交叉熵、MCE 和 M^3CE 进行对比实验，达到了预期的效果，充分证明本文方法的可行性，为今后的研究提供一种新的思路。

基于改进最小分类误差准则的深度学习算法，将非标签集中的梯度作区分处理导致批运行时间长，同时对特征简单的样本集的识别精度提高并不明显。因此，下一步研究并行化提高效率和进一步优化。

参考文献：

[1] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521 (7553): 436-444.

[2] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks [C]// Proc of European Conference on Computer Vision. Cham: Springer, 2014: 818-833.

[3] LeCun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86 (11): 2278-2324.

[4] Xiao Jianxiong, Hays J, Ehinger K A, *et al.* SUN database: large-scale scene recognition from abbey to zoo [C]// Proc of IEEE Computer Vision and Pattern Recognition. 2010: 3485-3492.

[5] 黄凯奇, 任伟强, 谭铁牛, 等 图像物体分类与检测算法综述 [J]. 计算机学报, 2014, 37 (6): 1225-1240. (Huang Kaiqi, Ren Weiqiang, Tan Tieniu. A review on image object classification and detection [J]. Chinese Journal of Computers, 2014, 37 (6): 1225-1240.)

[6] 何希平, 张琼华, 刘波. 基于 HOG 的目标分类特征深度学习模型 [J]. 计算机工程, 2016, 42 (12): 176-180. (He Xiping, Zhang Qionghua, Liu Bo. Deep learning model of target classification features based on HOG [J].

Computer Engineering, 2016, 42 (12): 176-80.)

[7] LeCun Y. Generalization and network design strategies [C]// Proc of Connectionism in Perspective. 1989.

[8] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]// Proc of International Conference on Neural Information Processing Systems. [S. l.] : Curran Associates Inc, 2012: 1097-1105.

[9] 余礼杨，范春晓．一种改进的多尺度核卷积目标跟踪算法 [J]．计算机应用，2015, (Xu Liyang, Fan Chun xiao. Multi-scale tracker based on improved kernel correlation filter [J]. Journal of Computer Application. 2015.)

[10] 贾涛，周利莉，陈健，等．基于卷积神经网络的 PCB CT 图像中的过孔和焊盘检测算法 [J]．计算机应用研究，2018, 35 (2): 637-640. (Jia Tao, Zhou Lili, Chen Jian, *et al.* Via and pad detection in PCB CT images based on convolutional neural network [J]. Application Research of Computer, 2018, 35 (2): 637-640.) .

[11] Finn C, Hendricks L A, Darrell T. Learning compact convolutional neural networks with nested dropout [J]. Eprint Arxiv, 2015,

[12] 归喆．基于深度学习的人脸特征提取与匹配 [D]；成都：电子科技大学，2016. (Gui Zhe. Facial feature extraction and matching based on deep learning [D]. Chengdu: University of Electronic Science and Technology of China, 2016.)

[13] Papoulis A, Saunders H. Probability, random variables and stochastic processes (2nd edition) [J]. A Papoulis, 1989, 33 (6): 1637-1637.

[14] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]// Proc of International Conference on Neural Information Processing Systems. [S. l.] : Curran Associates Inc, 2012: 1097-1105.

[15] Szegedy C, Wei Liu, Jia Yangqing, *et al.* Going deeper with convolutions [C]// Proc of IEEE Computer Vision and Pattern Recognition. 2015: 1-9.

[16] Juang B H, Katagiri S. Discriminative learning for minimum error classification [pattern recognition] [J]. IEEE Trans on Signal Processing, 1992, 40 (12): 3043-54.

[17] Feng Ziyong, Sun Zenghui, Jin Lianwen. Learning deep neural network using max-margin minimum classification error [C]// Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. , 2016.

[18] Bouvrie J. Notes on convolutional neural networks [J]. Neural Nets, 2006.